On measurement uncertainties derived from "Metrological Statistics"

Michael Grabe

Am Hasselteich 5, 38104 Braunschweig, Germany.
michael.grabe@ptb.de

Abstract

As measurement uncertainties are closely tied up with error models, it might be of interest to review a model, which the author assigns to "Metrological Statistics". Given that the random errors are normally distributed, the experimentalist could either refer to B.L. Welch's concept of "effective degrees of freedom" or to the multidimensional Fisher-Wishart distribution density. In the first case, different numbers of repeated measurements are admissible, in the latter it is strictly required to have equal numbers of repeated measurements. In error propagation, however, only the latter mode of action opens up the possibility of designing confidence intervals according to Student and confidence ellipsoids according to Hotelling. Another point of view, closely linked to the choice of the numbers of repeated measurements, refers to the customary practice of attributing equal rights to statistical expectations and empirical estimators. However, the Fisher-Wishart distribution density suggests using only the information which is realistically accessible to experimentalists, namely empirical estimators. For the handling of unknown systematic errors, either the existence of a (rectangular) distribution density may be assumed or, and this is proposed here, they may be classified as time-constant quantities, biasing expectations and suspending a lot of tools and procedures of error calculus well-established otherwise.

1 Introduction

The joint propagation of random errors and unknown systematic errors currently places the experimentalist in the following dilemma.

In regard to the propagation of *random errors*, there are, at least in principle, two different choices. If one is willing to accept *unequal numbers* of repeated measurements of the physical quantities to be combined within a given function, one has, in order to express the influence of random errors, to resort to B. L. Welch's sophisticated concept of so-called *numbers of effective degrees of freedom* [8]. However, this procedure is tied up with difficulties: it is restricted to independent variables.

Though B. L. Welch's concept completely exhausts the information implied in measured data, unfortunately, from a metrological point of view, it is cumbersome to handle and obstructs the view to existing simpler procedures. On the other hand, if the experimentalist preferred *equal numbers* of repeated measurements, he would — if need be — have to give away part of his information, namely that which is carried by the

excessive numbers of repeated measurements of the variables involved. Up to now, the disregarding of excessive numbers is regarded as unfavourable. In spite of this view, just this precaution opens up a toolbox of applied statistics hitherto closed to metrologists, as only with equal numbers of repeated measurements, is the experimentalist in a position to call upon the standard model of statistics for jointly normally distributed random variables, i.e. the Fisher-Wishart density [3]. The advantages gained in that way outweigh by far the "lost information", as relatively few repeated measurements of experimental set-ups, operating in a stationary mode, are able to locate accurately the respective physical quantities. After all, in error propagation the experimentalist may define confidence intervals according to Student (Gosset) including *any number* of variables. In least squares, he may even establish multidimensional confidence intervals, and last but not least, certain problems of classical error calculus, such as the Fisher–Behrens problems no longer arise.

In regard to the interpretation and propagation of *unknown systematic errors*, the situation is not simpler. Let us assume that an unknown systematic error f, constant in time, is confined to an interval of the kind[1]

$$-f_s \leq f \leq f_s, \quad f_s \geq 0. \tag{1.1}$$

Now, the experimentalist may either assign a postulated probabilty density to f, usually a rectangular density [7],

$$p(f) = \frac{1}{2f_s}, \tag{1.2}$$

or he may set without exception

$$f = \text{constant}, \tag{1.3}$$

where f lies anywhere within (1.1). The latter interpretation introduces biased estimators, leading to a break-down of many procedures of error calculus otherwise well-established.

Seen mathematically, both interpretations should be justified. In the case of (1.2), the combination of random and systematic errors should be carried out geometrically, in the case of (1.3), arithmetically. Regarding (1.3), the author suggests adding linearly Student's confidence intervals to appropriately designed worst-case estimates of the propagated systematic errors, and no probability statements should be associated with so-defined overall uncertainties.

2 Error propagation

The fundamental error equations of *Metrological Statistics* are given as follows [4]. Let x_0 designate *the true value* of the physical quantity x to be measured. Furthermore, let ε_l be the random error and f_x = constant the unknown systematic error corresponding

[1]Should the interval be unsymmetrical to zero, it could be symmetrized by subtracting the halved sum of the upper and lower boundary — the same quantity would have to be subtracted from the data.

to (1.1). We then have

$$x_l = x_0 + \varepsilon_l + f_x, \quad l = 1, \ldots, n. \tag{2.1}$$

Let $\mu_x = x_0 + f_x$ be the expectation of the random variable $X = \{x_1, x_2, \ldots, x_n\}$, so that the x_l are some of its realizations. We then find

$$x_l = \mu_x + \varepsilon_l, \quad l = 1, \ldots, n. \tag{2.2}$$

Furthermore, let $\bar{x} = 1/n \sum_{l=1}^{n} x_l$ denote the arithmetic mean. We then have the useful identities

$$x_l = x_0 + (x_l - \mu_x) + f_x, \qquad \bar{x} = x_0 + (\bar{x} - \mu_x) + f_x. \tag{2.3}$$

While the arithmetic mean is biased, the empirical variance

$$s_x^2 = \frac{1}{n-1} \sum_{l=1}^{n} (x_l - \bar{x})^2 \tag{2.4}$$

is not. For the time being, let us consider just two quantities to be measured, x and y. As robust and simple uncertainty assessments are *a matter of linearization*, the overall uncertainty u_ϕ of a given function $\phi(x, y)$ is proposed to be [5],

$$u_\phi = \frac{t_{S,P}(n-1)}{\sqrt{n}} \sqrt{\left(\frac{\partial \phi}{\partial \bar{x}}\right)^2 s_x^2 + 2\left(\frac{\partial \phi}{\partial \bar{x}}\right)\left(\frac{\partial \phi}{\partial \bar{y}}\right) s_{xy} + \left(\frac{\partial \phi}{\partial \bar{y}}\right)^2 s_y^2} + \left|\frac{\partial \phi}{\partial \bar{x}}\right| f_{s,x} + \left|\frac{\partial \phi}{\partial \bar{y}}\right| f_{s,y} \tag{2.5}$$

where $t_{S,P}(n-1)$ is the Student-factor corresponding to a confidence level P. We distinctly see how the empirical covariance

$$s_{xy} = \frac{1}{n-1} \sum_{l=1}^{n} (x_l - \bar{x})(y_l - \bar{y})$$

enters the empirical variance of the $\phi(x_l, y_l)$; $l = 1, \ldots, n$, given by

$$s_\phi^2 = \frac{1}{n-1} \sum_{l=1}^{n} [\phi(x_l, y_l) - \phi(\bar{x}, \bar{y})]^2 = \left(\frac{\partial \phi}{\partial \bar{x}}\right)^2 s_x^2 + 2\left(\frac{\partial \phi}{\partial \bar{x}}\right)\left(\frac{\partial \phi}{\partial \bar{y}}\right) s_{xy} + \left(\frac{\partial \phi}{\partial \bar{y}}\right)^2 s_y^2.$$

The final result

$$\phi(\bar{x}, \bar{y}) \pm u_\phi \tag{2.6}$$

is expected to localize the true value $\phi(x_0, y_0)$ with "reasonable certainty" — but no proper confidence statement should be added, as u_ϕ is a mixture of a statistical and a non statistical component. The last term in (2.5) may overestimate the uncertainty, on the other hand linearization errors have been negleted. After all, this uncertainty statement should fulfill the prerequiste to be safe, robust and simple.

If there are m quantities to be measured, we replace the notation $\bar{x}, \bar{y}$ by $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$. Then the overall uncertainty u_ϕ of the final result

$$\phi(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m) \pm u_\phi$$

is given by

$$u_\phi = \frac{t_{S,P}\,(n-1)}{\sqrt{n}} \sqrt{\sum_{i,j}^{m} \frac{\partial\phi}{\partial\bar{x}_i}\frac{\partial\phi}{\partial\bar{x}_j} s_{ij}} + \sum_{i=1}^{m} \left|\frac{\partial\phi}{\partial\bar{x}_i}\right| f_{s,i}. \tag{2.7}$$

When (2.5) and (2.7) are compared, it becomes obvious that the proposed formalism of error propagation works like a building kit, perspicuous and easy to handle. There are arguments against (2.7), in particular that an experimentalist who wishes to design his uncertainties in this way, would have to know *the complete set of repeated measurements*, in other words, the *complete* empirical variance-covariance matrix

$$s = (s_{ij}), \qquad i,j = 1,2,\ldots,m, \tag{2.8}$$

of the input data. Arguably, this is true, but in the days of computers and the internet such a challenge should no longer be apt to provoke difficulties worth mentioning. Another argument, that (2.7) might overestimate overall uncertainties, should be judged in view of the unique role of metrology in science. Standing "between" theory and experiment, metrology pursues the idea to localize reliably the value of the physical quantity in question.

3 Least squares

Let

$$A\beta \approx x \tag{3.1}$$

be a linear system of equations to be adjusted. Here, A designates the $m \times r$ design matrix of rank r, β the $r \times 1$ vector of unknowns and, finally, x the $m \times 1$ vector of the observations or input data. We assume $m > r$. The idea of least squares is of purely geometrical origin.

In what follows, A^T denotes the transpose of A. The idea is to project the vector x by means of a projection operator

$$P = A\left(A^T A\right)^{-1} A^T \tag{3.2}$$

orthogonally onto the column space of the matrix A, and the result is

$$\bar{\beta} = \left(A^T A\right)^{-1} A^T x. \tag{3.3}$$

As the solution vector $\bar{\beta}$ is linear in the input data, the transfer of (2.7) to its components $\bar{\beta}_k$, $k = 1,\ldots,r$, is straightforward.

Clearly, the orthogonal projection is in no way dependent on the error model implied. In contrast to this, the latter turns out to be crucial in regard to uncertainty assessments. Let us consider a set of *single observations*

$$x_i = x_{0,i} + \varepsilon_i + f_i = x_{0,i} + (x_i - \mu_i) + f_i, \quad i = 1,\ldots,m, \tag{3.4}$$

being the input data, where $E\{X_i\} = \mu_i$. Writing (3.4) in vector form, we have

$$x = x_0 + (x - \mu) + f \tag{3.5}$$

where

$$x = (x_1, x_2, \dots, x_m)^T, \qquad x_0 = (x_{0,1}, x_{0,2}, \dots, x_{0,m})^T,$$
$$\mu = (\mu_1, \mu_2, \dots, \mu_m)^T, \qquad f = (f_1, f_2, \dots, f_m)^T, \quad -f_{s,i} \le f_i \le f_{s,i}.$$

Given equal variances $\sigma^2 = E\left\{(X_i - \mu_i)^2\right\}$, the minimized sum $Q_{\min}$ of squared residuals of the adjusted system (3.1) should yield, according to quite familiar procedures, an estimator $s^2 \approx \sigma^2$. However, from

$$Q_{\min} = (x - Px)^T (x - Px),$$

we obtain something different, namely

$$E\{Q_{\min}\} = \sigma^2 (m - r) + f^T f - f^T P f. \tag{3.6}$$

As we see, even the simplest of all associated least squares procedures breaks down, should the model of time-constant unknown systematic errors be accepted. At the same time the related basic tool linked to $Q_{\min}$ and frequently used, namely *the test of consistency* of the input data based on the criterion

$$Q_{\min}/s^2 \approx m - r$$

breaks down as well. Indeed, during many decades, time and again, the observation

$$Q_{\min}/s^2 \gg m - r$$

has stunned experimentalists [2], so that, in the adjustments of the fundamental physical constants, even the abolition of least squares has been considered [1]. However, in view of (3.6), these observations are understandable.

After all, a least squares adjustment of *biased* input data requires arithmetic means

$$\bar{x}_i = x_{0,i} + (\bar{x}_i - \mu_i) + f_i, \quad i = 1, \dots, m, \tag{3.7}$$

so that the empirical variances and covariances

$$s_{ij} = \frac{1}{n-1} \sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j), \qquad s_{ii} = s_i^2, \tag{3.8}$$

are known *a priori*. Replacing (3.5) by

$$\bar{x} = x_0 + (\bar{x} - \mu) + f \tag{3.9}$$

instead of (3.3), we find

$$\bar{\beta} = \left(A^T A\right)^{-1} A^T \bar{x}. \tag{3.10}$$

A matter of similar concern refers to the break-down of the Gauss-Markoff theorem. In view of (3.9), the solution vector $\bar{\beta}$ is biased, so that the experimentalist is no longer in a position to obtain a weight-matrix from the variance-covariance matrix of the input vector $\bar{x}$. Consequently, simple, *optimized adjustments*, to which we are customarily used, must be ruled out. Nevertheless, we may multiply (3.1) from the left with any

non-singular weighting matrix, e.g. with a diagonal one,

$$G = \{g_1, g_2, \dots, g_m\}, \quad g_i = \frac{1}{u_{\bar{x}_i}}, \tag{3.11}$$

and adjust the weights g_i by trial and error in order to find the shortest possible uncertainty intervals. As has been shown, this method is also able to detect inconsistencies among the input data, [6]. Indeed, as a non-singular weight-matrix cannot shift the true solution vector β_0, we are allowed to proceed this way.

To assign uncertainties to the components $\bar{\beta}_k; k = 1, \dots, r$ of the solution vector $\bar{\beta}$, we refer to (2.7). To abbreviate the notation, we set in (3.10)

$$B = A\left(A^T A\right)^{-1} \tag{3.12}$$

where the elements of the matrix B will be designated by b_{ik}. Upon insertion of (3.9) into (3.10), we arrive at

$$\bar{\beta} = B^T x_0 + B^T(\bar{x} - \mu) + B^T f. \tag{3.13}$$

Evidently, $\beta_0 = B^T x_0$ is the true value of the estimator $\bar{\beta}$. Setting $\mu_{\bar{\beta}} = E\{\bar{\beta}\} = \beta_0 + B^T f$, we may define the *theoretical* variance-covariance matrix

$$E\left\{\left(\bar{\beta} - \mu_{\bar{\beta}}\right)\left(\bar{\beta} - \mu_{\bar{\beta}}\right)^T\right\},$$

which, however, remains numerically inaccessible. Consequently, the only thing we can do is to resort to the *empirical* variance-covariance matrix

$$s_{\bar{\beta}} = \left(s_{\bar{\beta}_k \bar{\beta}_{k'}}\right) = B^T s B, \quad k = 1, 2, \dots, r, \tag{3.14}$$

whose elements are given by

$$s_{\bar{\beta}_k \bar{\beta}_{k'}} = \sum_{i,j}^{m} b_{ik} b_{jk'} s_{ij}, \qquad s_{\bar{\beta}_k \bar{\beta}_k} = s^2_{\bar{\beta}_k}. \tag{3.15}$$

Clearly, the s_{ij} are the elements of the empirical variance-covariance matrix s of the input data, as has been stated in (2.8) and (3.8).

These procedures presuppose, as has been pointed out, *equal numbers* of repeated measurements within each of the m means (3.7). The components $\bar{\beta}_k$ of the solution vector may be written as

$$\bar{\beta}_k = \frac{1}{n}\sum_{l=1}^{n} \bar{\beta}_{kl} \quad \text{with} \quad \bar{\beta}_{kl} = \sum_{i=1}^{m} b_{ik} x_{il}; \quad k = 1, \dots, r. \tag{3.16}$$

Evidently, the $\bar{\beta}_{kl}$ are independent and normally distributed. Let $\mu_{\bar{\beta}_k}$ denote the expectations

$$\mu_{\bar{\beta}_k} = E\{\bar{\beta}_k\}, \quad k = 1, \dots, r \tag{3.17}$$

of the $\bar{\beta}_k$. Looking for just *any one* of the $\bar{\beta}_k$,

$$\bar{\beta}_k - \frac{t_{S,P}(n-1)}{\sqrt{n}} s_{\bar{\beta}_k} \le \mu_{\bar{\beta}_k} \le \bar{\beta}_k + \frac{t_{S,P}(n-1)}{\sqrt{n}} s_{\bar{\beta}_k} \tag{3.18}$$

is a confidence interval according to Student, where $t_{S,P}(n-1)$ is the Student-factor. This interval localizes $\mu_{\bar{\beta}_k}$ with confidence P.

The components of the third term on the right-hand side of (3.13) are given by

$$f_{\bar{\beta}_k} = \sum_{i=1}^{m} b_{ik} f_i, \quad k = 1, \ldots, r. \tag{3.19}$$

Worst-case estimates are

$$f_{s,\bar{\beta}_k} = \sum_{i=1}^{m} |b_{ik}| f_{s,i} l, \quad k = 1, \ldots, r. \tag{3.20}$$

After all, the overall uncertainties $u_{\bar{\beta}_k}$ of the components of the solution vector $\bar{\beta}$, considered and employed individually, are proposed to be

$$u_{\bar{\beta}_k} = \frac{t_{S,P}(n-1)}{\sqrt{n}} s_{\bar{\beta}_k} + f_{s,\bar{\beta}_k}, \quad k = 1, \ldots, r. \tag{3.21}$$

4 Uncertainty spaces

The component representation of (3.13),

$$\bar{\beta}_k = \beta_{0,k} + \sum_{i=1}^{m} b_{ik} (\bar{x}_i - \mu_i) + \sum_{i=1}^{m} b_{ik} f_i \tag{4.1}$$

reveals the couplings between the least squares estimators. Those due to random errors may be expressed by means of Hotelling's density [3]. The last term on the right-hand side of (4.1),

$$f_{\bar{\beta}_k} = \sum_{i=1}^{m} b_{ik} f_i, \quad k = 1, \ldots, r, \tag{4.2}$$

expresses the couplings due to systematic errors. The r components $f_{\bar{\beta}_k}$ map the m-dimensional hypercuboid

$$-f_{s,i} \leq f_i \leq f_{s,i}, \quad i = 1, \ldots, m, \tag{4.3}$$

onto the r-dimensional space, yielding a convex polytope. Both solids may be combined to an overall uncertainty space, resembling a "convex potato". Figures 1–3 show the confidence ellipsoid, the "security polytope" and the combination of both to an overall uncertainty space for the example of a least squares adjustment of a circle.

5 Conclusion

As computer simulations reveal, the approach presented here leads to measurement uncertainties safeguarding *physical objectivity* in the sense that uncertainty intervals reliably locate the values of the physical quantities in question. With such a distinct statement, the traceability of units and standards will certainly be maintained.

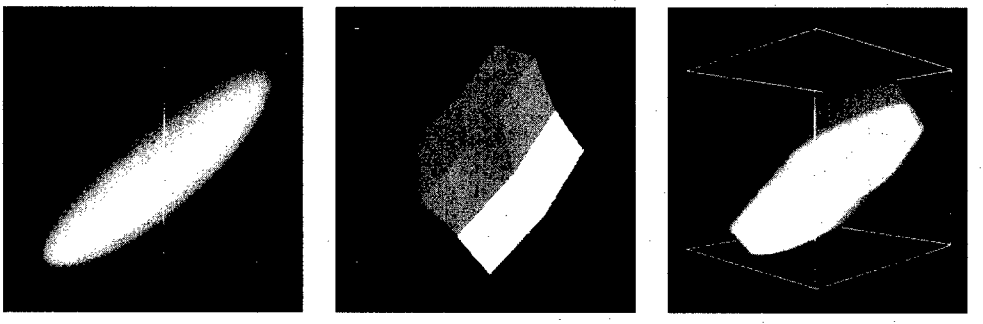

FIG. 1. Confidence ellipsoid, security polytope, overall uncertainty space resembling a "convex potato".

References

1. Bender, P.L., B. N. Taylor, E.R. Cohen, J.S. Thomas, P. Franken, and C. Eisenhart, *Should least squares adjustment of the fundamental constants be abolished?*, NBS Special Publication **343**, United States Department of Commerce, Washington D.C., 1971.
2. Cohen, E.R. and B.R. Taylor, *The 1986 adjustment of the fundamental physical constants*, CODATA BULLETIN Nr. **63** (1986).
3. Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton 1961.
4. Grabe, M., *Principles of "metrological statistics"*, metrologia **23** (1986/87) 213–219.
5. Grabe, M. *Estimation of measurement uncertainties, an alternative to the ISO-Guide*, metrologia **38** (2001) 97–106.
6. Grabe, M., An alternative algorithm for adjusting the fundamental physical constants, Physics Letters A **213** (1996) 125–137.
7. ISO, *Guide to the expression of uncertainty in measurement*, 1993. 1, Rue de Varambé, Boîte postale 56, CH-1211 Geneva 20, Switzerland.
8. Welch, B.L., *The generalization of Student's problem when several different population variances are involved*, Biometrika **34** (1947) 28–35.